Yu. E. Nesterov, A method of solving a convex programming problem with convergence rate $O\left(\frac{1}{k^2}\right)$, *Dokl. Akad. Nauk SSSR*, 1983, Volume 269, Number 3, 543–547

УДК 51 МАТЕМАТИКА

Ю.Е. НЕСТЕРОВ

МЕТОД РЕШЕНИЯ ЗАДАЧИ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ СО СКОРОСТЬЮ СХОДИМОСТИ $O\left(\frac{1}{k^2}\right)$

(Представлено академиком Л.В. Канторовичем 9 VII 1982)

1. В статье предлагается метод решения задачи выпуклого программирования в гильбертовом пространстве E. В отличие от большинства методов выпуклого программирования, предлагавшихся ранее, этот метод строит минимизирующую последовательность точек $\{x_k\}_{k=0}^{\infty}$, которая не является релаксационной. Эта особенность позволяет свести к минимуму вычислительные затраты на каждом шаге. В то же время для такого метода удается получить неулучшаемую на рассматриваемом классе задач оценку скорости сходимости (см. [1]).

2. Рассмотрим сначала задачу безусловной минимизации выпуклой функции $f(x)$. Мы будем предполагать, что функция $f(x)$ принадлежит классу $C^{1,1}(E)$, т.е. что существует константа $L > 0$, для которой при всех $x, y \in E$ выполняется неравенство

$$\|f'(x) - f'(y)\| \leqslant L\|x - y\|. \tag{1}$$

Из неравенства (1) следует, что при всех $x, y \in E$

$$f(y) \leqslant f(x) + \langle f'(x), y - x\rangle + 0{,}5L\|y - x\|^2. \tag{2}$$

Для решения задачи $\min\{f(x) \mid x \in E\}$ с непустым множеством минимумов X^* предлагается следующий метод.

0) Выбираем точку $y_0 \in E$. Полагаем

$$k = 0, \quad a_0 = 1, \quad x_{-1} = y_0, \quad \alpha_{-1} = \|y_0 - z\| / \|f'(y_0) - f'(z)\|, \tag{3}$$

где z – любая точка из $E, z \neq y_0\ f'(z) \neq f'(y_0)$.

1) k-я Итерация.

а) Вычисляем наименьший номер $i \geqslant 0$, для которого выполняется неравенство

$$f(y_k) - f(y_k - 2^{-i}\alpha_{k-1} f'(y_k)) \geqslant 2^{-i-1}\alpha_{k-1}\|f'(y_k)\|^2. \tag{4}$$

б) Полагаем

$$\begin{gathered} \alpha_k = 2^{-i}\alpha_{k-1}, \ x_k = y_k - \alpha_k f'(y_k), \\ a_{k+1} = (1 + \sqrt{4a_k^2 + 1})/2, \\ y_{k+1} = x_k + (a_k - 1)(x_k - x_{k-1})/a_{k+1}. \end{gathered} \tag{5}$$

Способ прерывания одномерного поиска (4) аналогичен способу, предложенному в [2]. Разница лишь в том, что в (4) дробление шага на k-й итерации производится, начиная с α_{k-1} (а не с единицы, как в [2]). В силу этого (см. доказательство теоремы 1) при построении методом (3)–(5) последовательности $\{x_k\}_{k=0}^{\infty}$ будет сделано не более $O(\log_2 L)$ таких дроблений. Пересчет точек y_k в (5) осуществляется с помощью "овражного" шага. Отметим также, что метод (3)–(5) не обеспечивает монотонное убывание функции $f(x)$ на последовательностях $\{x_k\}_{k=0}^{\infty}$, $\{y_k\}_{k=0}^{\infty}$.

Теорема 1. *Пусть выпуклая функция $f(x) \in C^{1,1}(E)$ и $X^* \neq \phi$. Если последовательность $\{x_k\}_{k=0}^{\infty}$ построена методом* (3)–(5), *то*:

1) *для любого* $k \geqslant 0$

$$f(x_k) - f^* \leqslant C/(k+2)^2, \tag{6}$$

где $C = 4L\|y_0 - x^*\|^2$, $f^* = f(x^*)$, $x^* \in X^*$;

2) *для достижения точности* ϵ *по функционалу необходимо*:

а) *вычислить градиент целевой функции не более* $NG =]\sqrt{C/\epsilon}[$ *раз,*

б) *вычислить значение целевой функции не более* $NF = 2NG +]\log_2(2L\alpha_{-1})[+1$ *раз.*

Здесь и далее $]\,(\cdot)\,[$ – целая часть числа $(\cdot)$.

Доказательство. Пусть $y_k(\alpha) = y_k - \alpha f'(y_k)$. Из неравенства (2) получаем $f(y_k) - f(y_k(\alpha)) \geqslant 0{,}5\alpha(2 - \alpha L)\|f'(y_k)\|^2$. Следовательно, как только $2^{-i}\alpha_{k-1}$ станет меньше, чем L^{-1}, неравенство (4) выполнится и в дальнейшем α_k уменьшаться не будут. Таким образом, $\alpha_k \geqslant 0{,}5L^{-1}$ для всех $k \geqslant 0$.

Обозначим $p_k = (a_k - 1)(x_{k-1} - x_k)$. Тогда $p_{k+1} - x_{k+1} = p_k - x_k + a_{k+1}\alpha_{k+1}f'(y_{k+1})$. Следовательно, $\|p_{k+1} - x_{k+1} + x^*\|^2 = \|p_k - x_k + x^*\|^2 + 2(a_{k+1} - 1)\alpha_{k+1}\langle f'(y_{k+1}), p_k\rangle + 2a_{k+1}\alpha_{k+1}\langle f'(y_{k+1}), x^* - y_{k+1}\rangle + a_{k+1}^2\alpha_{k+1}^2 \times \|f'(y_{k+1})\|^2$.

Пользуясь неравенством (4) и выпуклостью функции $f(x)$, получаем

$$\langle f'(y_{k+1}), y_{k+1} - x^*\rangle \geqslant f(x_{k+1}) - f^* + 0{,}5\alpha_{k+1}\|f'(y_{k+1})\|^2,$$

$$0{,}5\alpha_{k+1}\|f'(y_{k+1})\|^2 \leqslant f(y_{k+1}) - f(x_{k+1}) \leqslant f(x_k) - f(x_{k+1}) - a_{k+1}^{-1}\langle f'(y_{k+1}), p_k\rangle.$$

Подставим эти два неравенства в предыдущее равенство:

$$\begin{aligned}&\|p_{k+1} - x_{k+1} + x^*\|^2 - \|p_k - x_k + x^*\|^2 \leqslant 2(a_{k+1} - 1)\alpha_{k+1}\langle f'(y_{k+1}), p_k\rangle - \\ &-2a_{k+1}\alpha_{k+1}(f(x_{k+1} - f^*) + (a_{k+1}^2 - a_{k+1})\alpha_{k+1}^2\|f'(y_{k+1})\|^2 \leqslant \\ &\leqslant -2a_{k+1}\alpha_{k+1}(f(x_{k+1}) - f^*) + 2(a_{k+1}^2 - a_{k+1})\alpha_{k+1}(f(x_k) - f(x_{k+1})) = \\ &= 2\alpha_{k+1}a_k^2(f(x_k) - f^*) - 2\alpha_{k+1}a_{k+1}^2(f(x_{k+1}) - f^*) \leqslant 2\alpha_k a_k^2(f(x_k) - f^*) - \\ &- 2\alpha_{k+1}a_{k+1}^2(f(x_{k+1}) - f^*).\end{aligned}$$

Таким образом,

$$\begin{aligned}&2\alpha_{k+1}a_{k+1}^2(f(x_{k+1}) - f^*) \leqslant 2\alpha_{k+1}a_{k+1}^2(f(x_{k+1}) - f^*) + \\ &+ \|p_{k+1} - x_{k+1} + x^*\|^2 \leqslant 2\alpha_k a_k(f(x_k) - f^*) + \|p_k - x_k + x^*\|^2 \leqslant \\ &\leqslant 2\alpha_0 a_0^2(f(x_0) - f^*) + \|p_0 - x_0 + x^*\|^2 \leqslant \|y_0 - x^*\|^2.\end{aligned}$$

Осталось заметить, что $a_{k+1} \geqslant a_k + 0{,}5 \geqslant 1 + 0{,}5(k+1)$.

Из оценки скорости сходимости (6) следует, что число итераций, необходимое методу (3)–(5) для достижения точности ϵ, не будет больше, чем $]\sqrt{C/\epsilon}[\, - 1$. При этом на каждой итерации будет вычисляться один градиент и по крайней мере два значения целевой функции. Заметим, однако, что каждому дополнительному вычислению значения целевой функции соответствует уменьшение величины α_k вдвое. Поэтому общее число таких вычислений не превзойдет $]\log_2(2L\alpha_{-1})[+1$. Теорема доказана.

Если для градиента целевой функции известна константа Липшица L, то в методе (3)–(5) можно положить $\alpha_k \equiv L^{-1}$ при любом $k \geqslant 0$. В этом случае неравенство (4) будет заведомо выполнено и поэтому утверждения теоремы 1 останутся верными при $C = 2L\|y_0 - x^*\|^2$, $NG =]\,\|y_0 - x^*\|\sqrt{2L/\epsilon}[\, - 1$ и $NF = 0$.

В заключение этого раздела покажем, как можно модифицировать метод (3) – (5) для решения задачи минимизации сильно выпуклой функции.

Предположим, что для функции $f(x)$ при всех $x \in E$ выполняется неравенство $f(x) - f^* \geqslant 0{,}5m\|x - x^*\|^2$, где $m > 0$, и пусть константа m нам известна.

Введем в метод (3) – (5) следующее правило прерывания:

в) Останавливаемся, если

(7) $\quad k \geqslant 2\sqrt{2/(m\alpha_k)} - 2.$

Пусть прерывание произошло на N-м шаге. Так как в методе (3) – (5) $\alpha_k \geqslant 0{,}5L^{-1}$, то $N \leqslant\ ]4\sqrt{L/m}[\ - 1$. В то же время

$$f(x_N) - f^* \leqslant \frac{2\|y_0 - x^*\|^2}{\alpha_N(N+2)^2} \leqslant 0{,}25m\|y_0 - x^*\|^2 \leqslant 0{,}5(f(y_0) - f^*).$$

После того как получена точка x_N, необходимо обновить метод и опять начать счет методом (3) – (5), (7) из точки x_N как из начальной и т.д.

В результате получаем, что за каждые $]4\sqrt{L/m}[\ - 1$ итераций невязка по функции убывает вдвое. Таким образом, метод (3) – (5) с обновлением (7) является неулучшаемым (с точностью до безразмерной константы) среди методов первого порядка на классе сильно выпуклых функций из $C^{1,1}(E)$ (см. [1]).

3. Рассмотрим следующую экстремальную задачу:

(8) $\quad \min\{F(\bar f(x)) \mid x \in Q\},$

где Q – выпуклое замкнутое множество из E, $F(u)$, $u \in R^m$, – выпуклая на всем R^m положительно-однородная степени единица функция, $\bar f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$ – вектор выпуклых непрерывно дифференцируемых на E функций. Множество X^* решений задачи (8) всегда предполагается непустым. Кроме того, мы всегда будем предполагать, что система функций $\{F(\cdot), \bar f(\cdot)\}$ обладает следующим свойством:

(*) Если существует вектор $\lambda \in \partial F(0)$ такой, что $\lambda^{(k)} < 0$, то $f_k(x)$ – линейная функция.

Через $\partial F(0)$ в (*) обозначен субдифференциал функции $F(u)$ в нуле.

Как известно, для выпуклых положительно-однородных степени единица функций справедливо тождество $F(u) \equiv \max\{\langle\lambda, u\rangle \mid \lambda \in \partial F(0)\}$. Поэтому из предположения (*) следует выпуклость функции $F(\bar f(x))$ на всем E.

Задачу (8) можно записать в минимаксной форме:

(9) $\quad \min\{\max\{\langle\lambda, \bar f(x)\rangle \mid \lambda \in \partial F(0)\} \mid x \in Q\}.$

Можно показать, что из непустоты множества X^* и предположения (*) следует существование у задачи (9) седловой точки (λ^*, x^*). Поэтому множество седловых точек задачи (9) представимо в виде $\Omega^* = \Lambda^* \times X^*$, где $\Lambda^* = \operatorname{Arg\,max}\{\Psi(\lambda) \mid \lambda \in \partial F(0)\}$, $\Psi(\lambda) = \min\{\langle\lambda, f(x)\rangle \mid x \in Q\}$. Задачу

$$\max\{\Psi(\lambda) \mid \lambda \in \partial F(0) \cap \operatorname{dom}\Psi(\cdot)\}.$$

мы будем называть задачей, двойственной к (8).

Пусть в задаче (8) функции $f_k(x)$, $k = 1, 2, \ldots, m$, принадлежат, классу $C^{1,1}(E)$ с константами $L^{(k)} \geqslant 0$. Обозначим $\bar L = (L^{(1)}, L^{(2)}, \ldots, L^{(m)})$.

Рассмотрим функцию $\Phi(y, A, z) = F(\bar f(y, z)) + 0{,}5A\|y - z\|^2$, где $\bar f(y, z) = (f^{(1)}(y, z), f^{(2)}(y, z), \ldots, f^{(m)}(y, z))$, $f^{(k)}(y, z) = f_k(y) + \langle f'(y), z - y\rangle$, $k = 1, 2, \ldots, m$, A – положительная константа. Обозначим

$$\Phi^*(y, A) = \min\{\Phi(y, A, z) \mid z \in Q\}, \qquad T(y, A) = \operatorname{argmin}\{\Phi(y, A, z) \mid z \in Q\}.$$

Отметим, что отображение $y \to T(y, A)$ является естественным обобщением для задачи (8) "градиентного" отображения, введенного в [1] в связи с исследованием методов минимизации функций вида $\max\limits_{1 \leqslant k \leqslant m} f_k(x)$. Для отображения $y \to T(y, A)$ (как и для "градиентного" отображения" из [1]) при всех $x \in Q$, $y \in E$, $A \geqslant 0$ выполняется неравенство

$$(10) \quad \Phi^*(y, A) + A\langle y - T(y, A), x - y\rangle + 0{,}5A\|y - T(y, A)\|^2 \leqslant F(\bar f(x)),$$

причем если $A \geqslant F(L)$, то

$$\Phi^*(y, A) \geqslant F(\bar f(T(y, A))).$$

Для решения задачи (8) предлагается следующий метод.

0) Выбираем точку $y_0 \in E$. Полагаем

$$(11) \quad k = 0, \quad a_0 = 1, \quad x_{-1} = y_0, \quad A_{-1} = F(\bar L_0),$$

где $\bar L_0 = (L_0^{(1)}, L_0^{(2)}, \ldots, L_0^{(m)})$, $L_0^{(k)} = \|f'_k(y_0) - f'_k(z)\| / \|y_0 - z\|$, z — произвольная точка из E, $z \neq y_0$.

1) k-я Итерация.

а) Вычисляем наименьший номер $i \geqslant 0$, для которого выполняется неравенство

$$(12) \quad \Phi^*(y_k, 2^i A_{k-1}) \geqslant F(\bar f(T(y_k, 2^i A_{k-1}))).$$

б) Полагаем $A_k = 2^i A_{k-1}$, $x_k = T(y_k, A_k)$,

$$(13) \quad \begin{aligned} a_{k+1} &= (1 + \sqrt{4a_k^2 + 1})/2, \\ y_{k+1} &= x_k + (a_k - 1)(x_k - x_{k-1})/a_{k+1}. \end{aligned}$$

Нетрудно заметить, что метод (3)—(5) является просто другой формой записи метода (11)—(13) для задачи безусловной минимизации (т.е. когда в (8) $m = 1$, $F(y) = y$, $Q = E$).

Теорема 2. *Если последовательность* $\{x_k\}_{k=0}^{\infty}$ *построена методом* (11)—(13), *то*:

1) *для любого* $k \geqslant 0$ $F(\bar f(x_k)) - F(\bar f(x^*)) \leqslant C_1/(k + 2)^2$, *где* $C_1 = 4F(\bar L)\|y_0 - x^*\|^2$, $x^* \in X^*$.

2) *для достижения точности* ϵ *по функционалу необходимо*:

а) *решить вспомогательную задачу* $\min\{\Phi(y_k, A, x) \mid x \in Q\}$ *не более* $]\sqrt{C_1/\epsilon}[+] \max\{\log_2(F(\bar L)/A_{-1}), 0\}[$ *раз,*

б) *вычислить набор градиентов* $f'_1(y), f'_2(y), \ldots, f'_m(y)$ *не более* $]\sqrt{C_1/\epsilon}[$ *раз,*

в) *вычислить вектор-функцию* $\bar f(x)$ *не более* $2]\sqrt{C_1/\epsilon}[+]\max\{\log_2(F(\bar L)/A_{-1}), 0\}[$ *раз.*

Теорема 2 доказывается практически так же, как и теорема 1. Необходимо только вместо неравенства (2) использовать неравенство (10), при этом аналогом вектора $\alpha_k f'(y_k)$ будет вектор $y_k - T(y_k, A_k)$, а аналогом α_k — величины A_k^{-1}.

Точно так же, как и в методе (3)—(5), в методе (11)—(13) можно учесть информацию о константе $F(\bar L)$ и параметре сильной выпуклости функции $F(\bar f(x))$ — m (для этого, правда, необходимо, чтобы $y_0 \in Q$).

В заключение отметим два важных частных случая задачи (8), в которых вспомогательная задача $\min\{\Phi(y_k, A, x) \mid x \in Q\}$ оказывается достаточно простой.

а) Минимизация гладкой выпуклой функции на простом множестве. Под простым множеством мы понимаем такое множество, для которого оператор проектирования записывается в явном виде. В этом случае в задаче (8) $m = 1$, $F(y) = y$

и в методе (11) – (13)

$$\Phi^*(y, A) = f(y) - 0{,}5A^{-1}\|f'(y)\|^2 + 0{,}5A\|T(y, A) - y + A^{-1}f'(y)\|^2,$$

где $T(y, A) = \operatorname{argmin}\{\|y - A^{-1}f'(y) - z\| \mid z \in Q\}$.

б) Безусловная минимизация (в задаче (8) $Q \equiv E$). В этом случае вспомогательная задача $\min\{\Phi(y, A, x) \mid x \in E\}$ эквивалентна следующей двойственной задаче:

$$(14)\quad \max\left\{-0{,}5A^{-1}\left\|\sum_{k=1}^{m}\lambda^{(k)}f'_k(y)\right\|^2 + \sum_{k=1}^{m}\lambda^{(k)}f_k(y) \mid (\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(m)}) \in \in \partial F(0)\right\}.$$

При этом $T(y, A) = y - A^{-1}\sum_{k=1}^{m}\lambda^{(k)}(y)f'_k(y)$, где $\lambda^{(k)}(y)$, $k = 1, 2, \ldots, m$, – решения задачи (14) при фиксированном $y \in E$. Отметим, что множество $\partial F(0)$ обычно задается простыми ограничениями – линейными либо квадратичными. В таких случаях задача (14) – стандартная задача квадратичного программирования.

Автор искренне признателен А.С. Немировскому за беседы, которые стимулировали его интерес к рассмотренным вопросам.

Центральный экономико-математический институт
Академии наук СССР, Москва

Поступило
19 VII 1982

ЛИТЕРАТУРА

1. *Немировский А.С., Юдин Д.Б.* Сложность задач и эффективность методов оптимизации. М.: Наука, 1979. 2. *Пшеничный Б.Н., Данилин Ю.М.* Численные методы в экстремальных задачах. М.: Наука, 1975.

УДК 515.1 МАТЕМАТИКА

Е.И. НОЧКА

К ТЕОРИИ МЕРОМОРФНЫХ КРИВЫХ

(Представлено академиком В.С. Владимировым 18 V 1982)

1. Пусть задана мероморфная кривая, т.е. мероморфное отображение

$$\tilde{f}\colon \quad \mathbf{C} \to \mathbf{CP}^n,$$

и пусть голоморфное отображение

$$f\colon \ \mathbf{C} \to \mathbf{C}^{n+1}, \qquad f = (f_1, f_2, \ldots, f_{n+1}),$$

является редуцированным представлением кривой $\tilde{f}$. Характеристическую функцию $\tilde{f}$ определим, следуя А. Картану [1]:

$$T(\tilde{f}, r) = \frac{1}{2\pi}\int_0^{2\pi}\log|f(re^{i\gamma})|^2 d\gamma - \log|f(0)|^2.$$

Пусть A – гиперплоскость в $\mathbf{CP}^n$ и a – единичный вектор такой, что равенство $(w, a) = 0$ (скобки обозначают эрмитово скалярное произведение) есть уравнение гиперплоскости A в однородных координатах; обозначим $f_A = (f, a)$.